Data Capstone Project: Analyzing New York School District Data

Yiqing Guo, Tavianne Kemp, Stanley Perez, Paul Polsinelli

Our Research Process: An Overview

1. Brainstormed draft questions in the world of school funding
2. Narrowed our scope from worldwide to the district level in New York state
3. Data source collection to answer our questions
4. General exploratory data analysis
5. Solidification of our questions
6. ETL process
7. Visualization creation
8. Machine Learning

Our Data Sources

01

NYSED Website

The NYSED website was scraped for school district performance, demographics, and per-student expenditure on a district level

02

NYC Open Data and 2019 NY State District Data

Provided information on funding sources for NY school districts

03

U.S. Census Bureau

Provided median income by zip code which was used to find median income by district

04

NCES School District Geographic Relationship Files

Used to map zip codes to school districts.

Exploratory Data Analysis

87.92%
Average Graduation Rate

$77,834.90
Average Median Household Income

27.09%
Minority Students

$23,251.60
Average Funding Received Per Student

2,420,688
Students enrolled in 2019

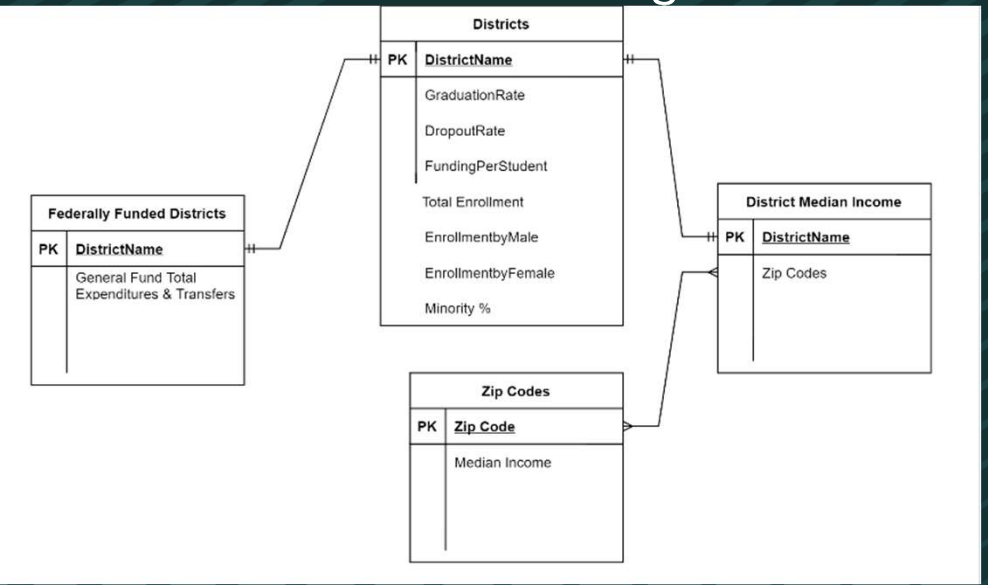
4.77%
Average Dropout Rate

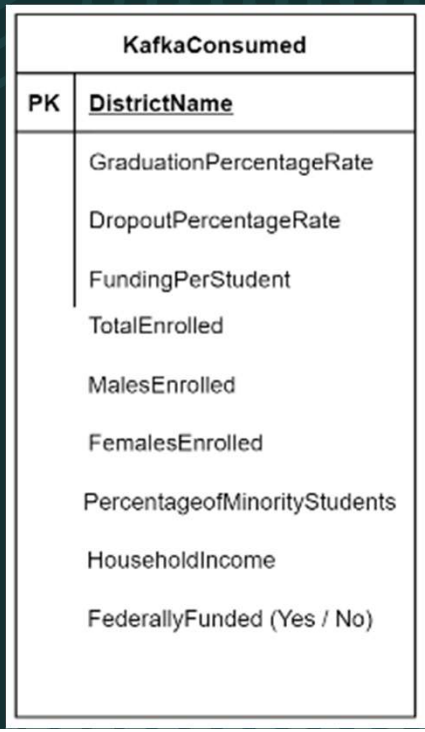
Exploratory Questions

- Which districts have the most funding per student and what differentiates them from the other districts?
- What percentage of districts receive federal funding and how do they compare to the ones that did not?
- What kind of correlation is there between the graduation rate and the other factors we've acquired?
- Is there a discrepancy between the graduation rate and the percentage of minority students in a district?
- Can we predict the graduation rate for a district based on the data and which factor has the most effect on the graduation rate for a school district?

ETL Process

1. Extracted CSV's from our data sources. Scraped the NYSED website
2. Delete unnecessary columns
3. Transform the median income data from zip codes to school districts
4. Rename some school districts to match other tables
5. Dump each table in a SQL server before being later combined
    1. Left ER diagram is our before, and right is the after





ETL Process for the School District Map

1. NYC school district mapping data created manually combining NYC zip code GIS file and NYC school districts GIS file
2. Extract map data from NYS GIS dataset

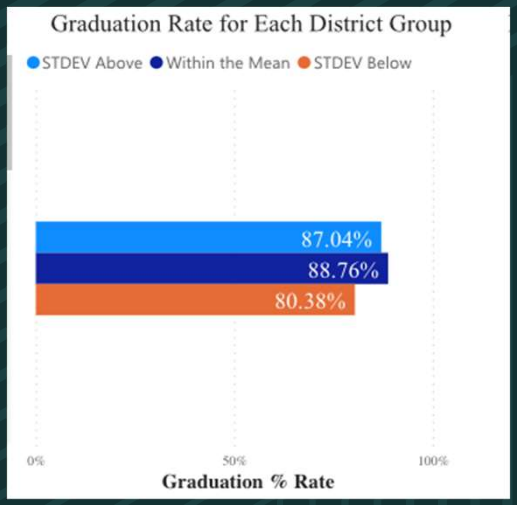
Which districts have the most funding per student and what differentiates them from the other districts? Is there a difference in graduation rate for these districts compared to the ones that don't receive as much funding?

Mean: $23,251

Districts high in funding: > $32,872

Remaining Districts: $13,630 > x < $32,872

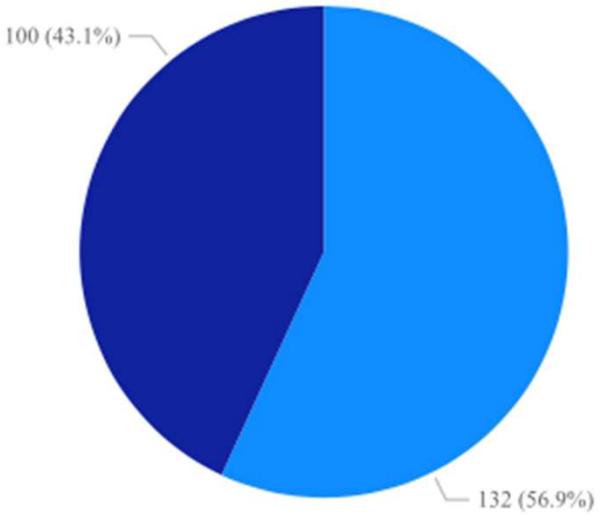
Districts low in funding: < $13,630

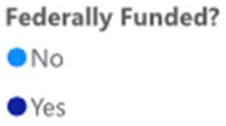


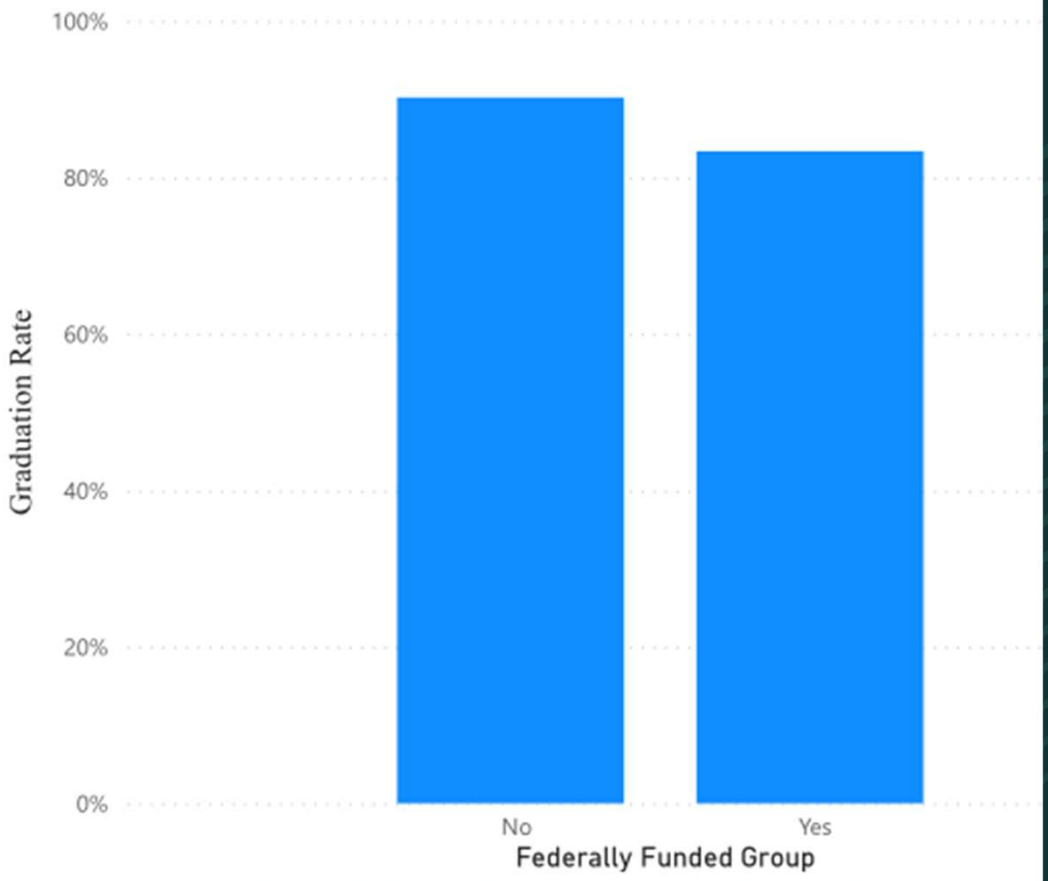
	GraduationPercentageRate	DropoutPercentageRate	FundingPerStudent	TotalEnrolled	MalesEnrolled	FemalesEnrolled	Minority	HouseholdIncome
count	49.000000	49.000000	49.000000	49.000000	49.000000	49.000000	49.000000	47.000000
mean	87.571429	3.979592	48404.679184	1032.102041	541.387755	490.714286	32.085714	89658.304681

	GraduationPercentageRate	DropoutPercentageRate	FundingPerStudent	TotalEnrolled	MalesEnrolled	FemalesEnrolled	Minority	HouseholdIncome
count	14.000000	14.000000	14.000000	14.000000	14.000000	14.00000	14.000000	13.000000
mean	81.285714	6.142857	9386.046429	27535.928571	14570.928571	12965.00000	66.907143	68659.110000

What is the percentage of districts that get federal funding versus ones that do not? How do the districts that received federal funding compare to the ones that didn't?







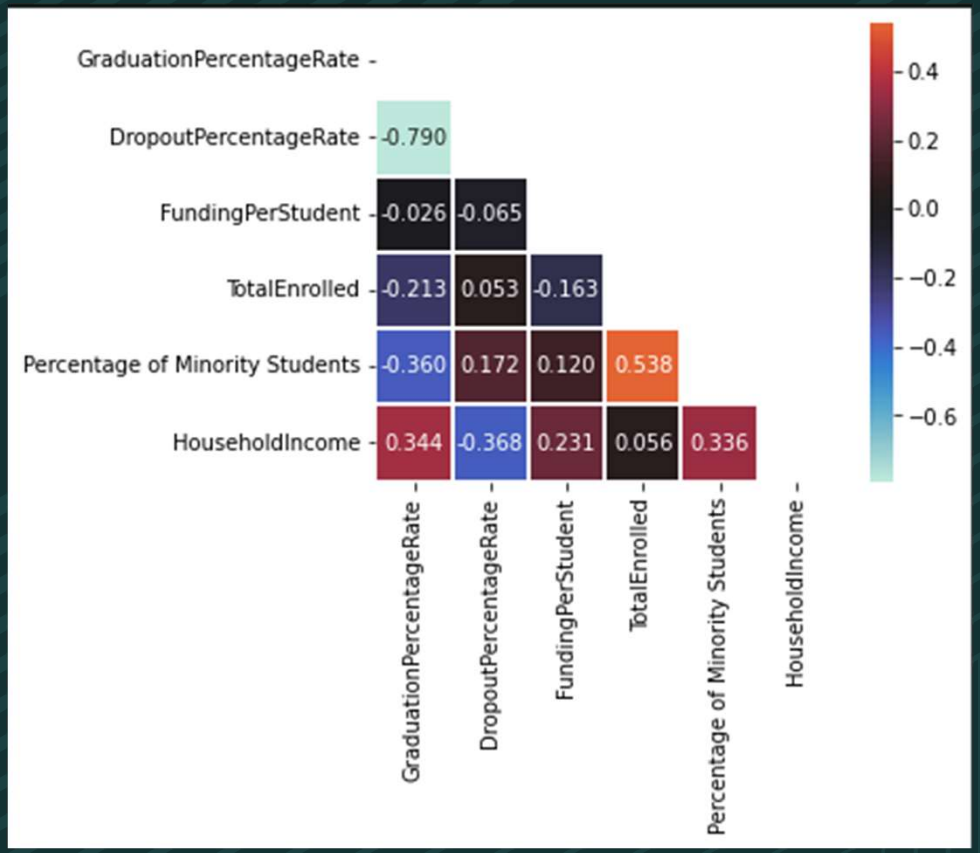
What is the percentage of districts that get federal funding versus ones that do not? How do the districts that received federal funding compare to the ones that didn't?

Federally funded districts -

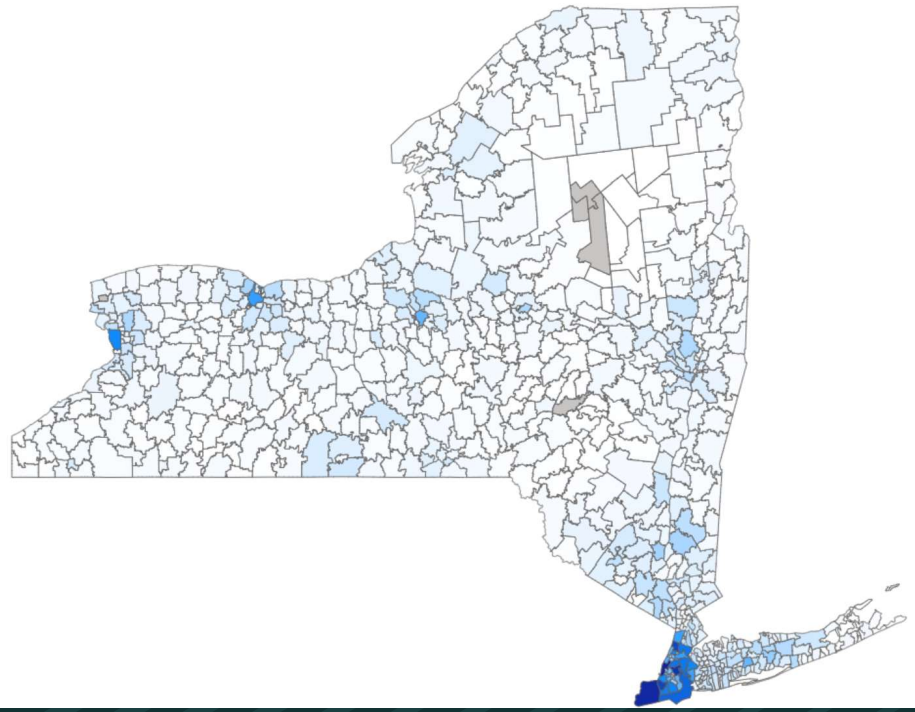
	GraduationPercentageRate	DropoutPercentageRate	FundingPerStudent	TotalEnrolled	MalesEnrolled	FemalesEnrolled	Minority	HouseholdIncome
count	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	97.000000
mean	83.360000	6.360000	21122.053500	11889.160000	6140.020000	5749.140000	50.450000	68754.330515

Non federally funded districts -

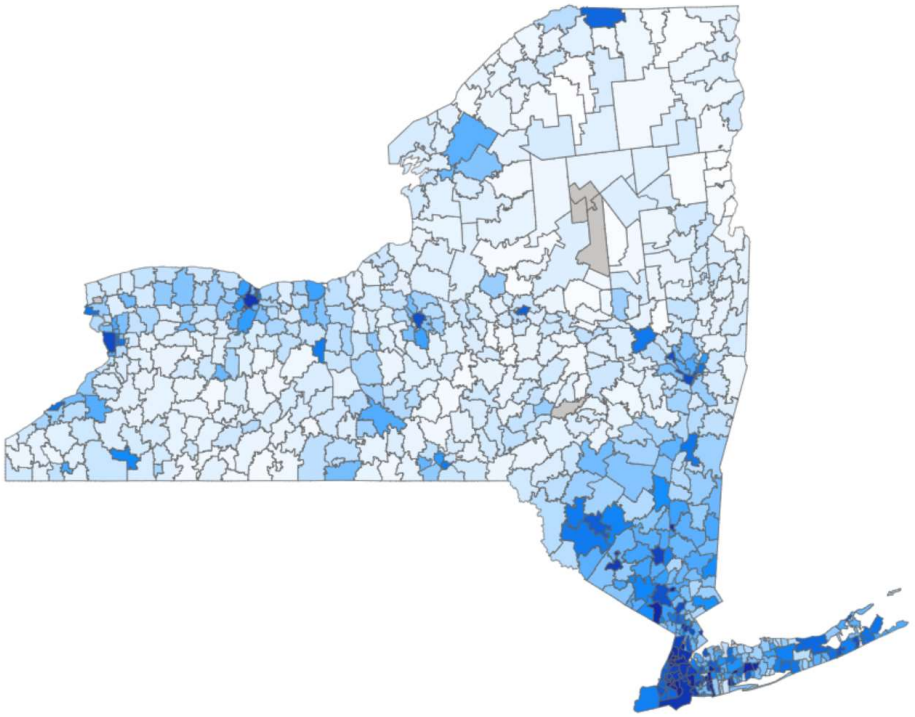
	GraduationPercentageRate	DropoutPercentageRate	FundingPerStudent	TotalEnrolled	MalesEnrolled	FemalesEnrolled	Minority	HouseholdIncome
count	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	132.000000	131.000000
mean	90.227273	3.765152	21962.976515	4104.393939	2103.121212	2001.272727	35.712121	95075.407481

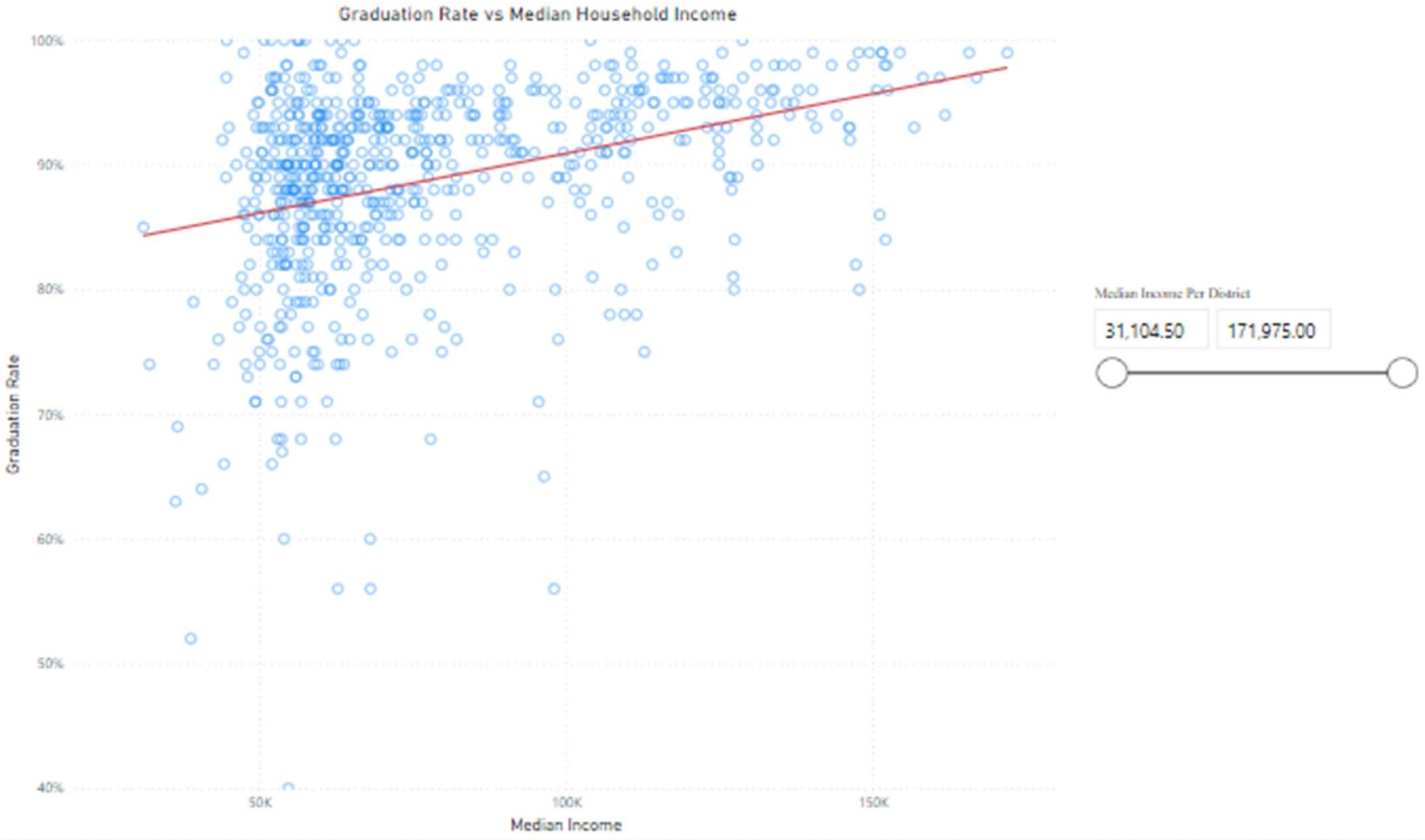

GraduationPercentageRate
DropoutPercentageRate
FundingPerStudent
TotalEnrolled
Percentage of Minority Students
HouseholdIncome
-0.790
-0.026
-0.065
-0.213
0.053
-0.163
-0.360
0.172
0.120
0.538
0.344
-0.368
0.231
0.056
0.336
0.4
0.2
0.0
−0.2
−0.4
−0.6

Total Enrollment by District

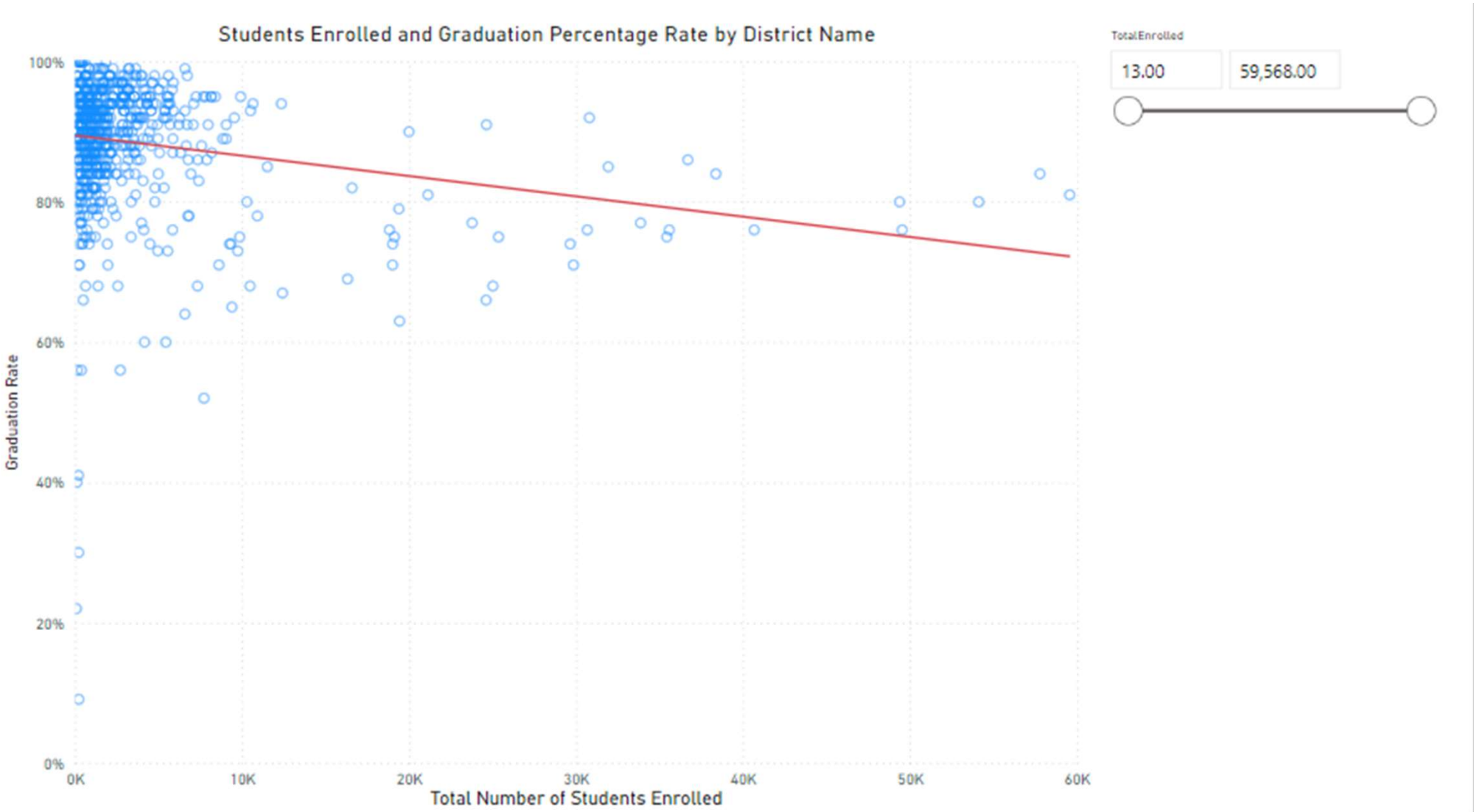


Percentage of Minority Students by District



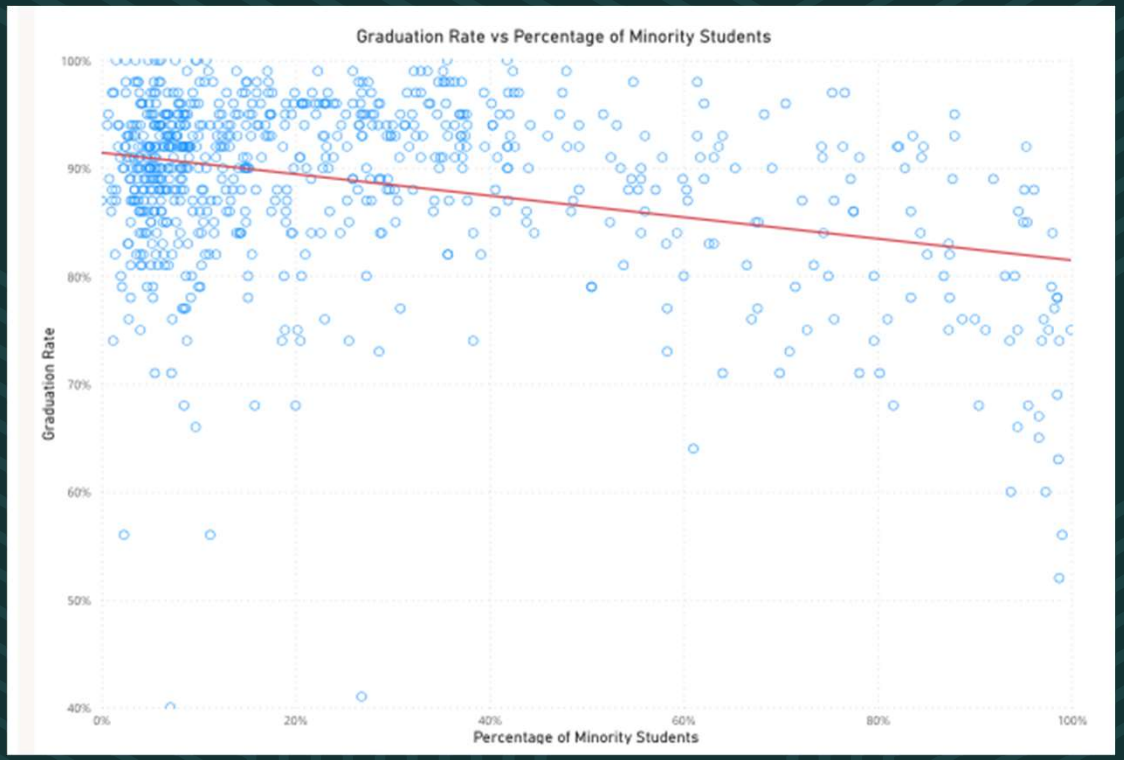


Is there a correlation between graduation rate and median income?

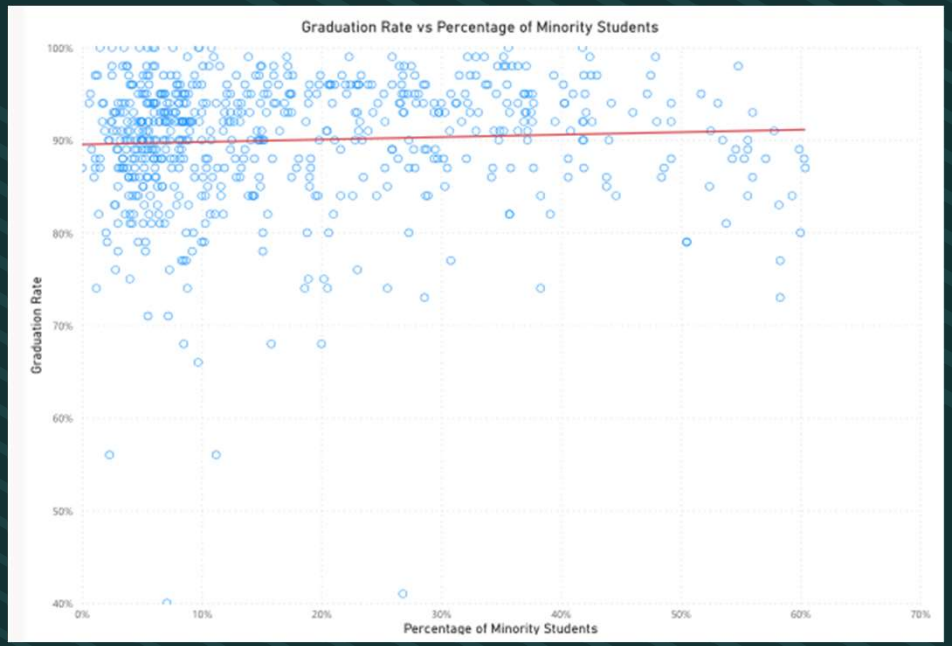


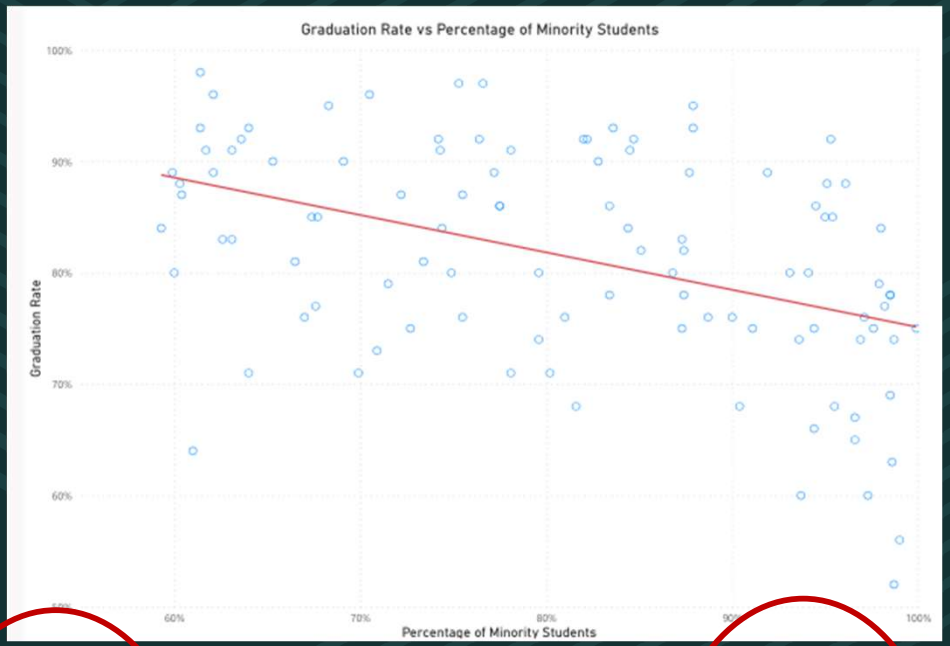
Is there a correlation between graduation rate and number of students enrolled?

Is there a discrepancy between the graduation rate and the percentage of minority students in a district?



Is there a discrepancy between the graduation rate and the percentage of minority students in a district?



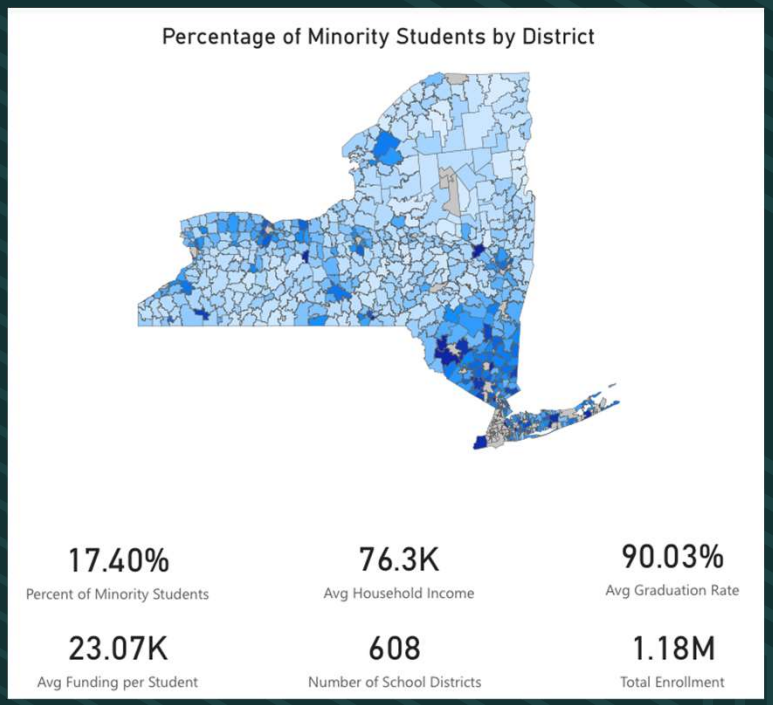


	GraduationPercentageRate	DropoutPercentageRate	FundingPerStudent	TotalEnrolled	MalesEnrolled	FemalesEnrolled	Minority	HouseholdIncome
count	609.000000	609.000000	609.000000	609.000000	609.000000	609.000000	609.000000	585.000000
mean	90.031199	4.338259	23066.290805	1954.374384	999.090312	955.284072	17.470115	76361.012427

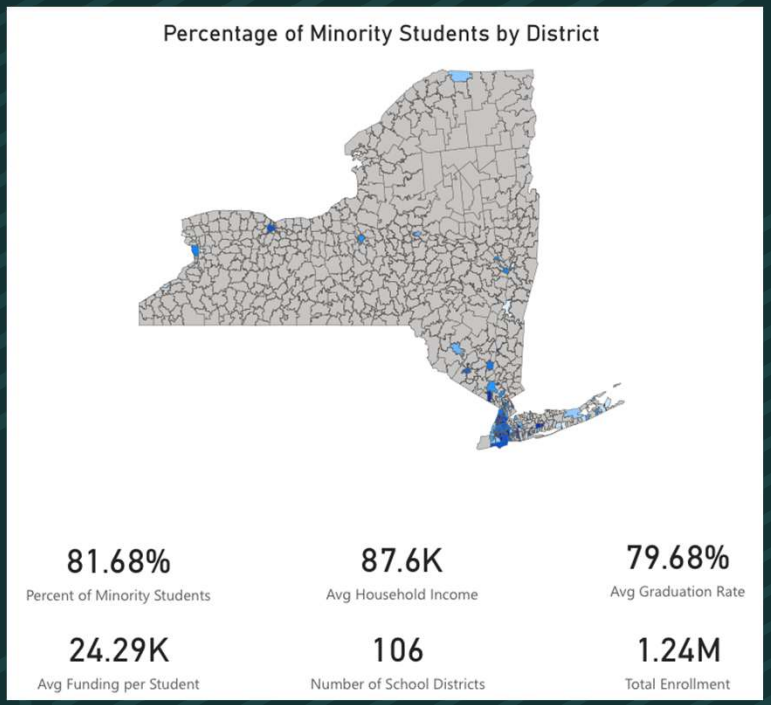
	GraduationPercentageRate	DropoutPercentageRate	FundingPerStudent	TotalEnrolled	MalesEnrolled	FemalesEnrolled	Minority	HouseholdIncome
count	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	104.000000	91.000000
mean	79.586538	7.307692	24327.020577	11815.932692	6110.000000	5705.932692	82.100000	87733.820769

Is there a discrepancy between the graduation rate and the percentage of minority students in a district?

< 60%

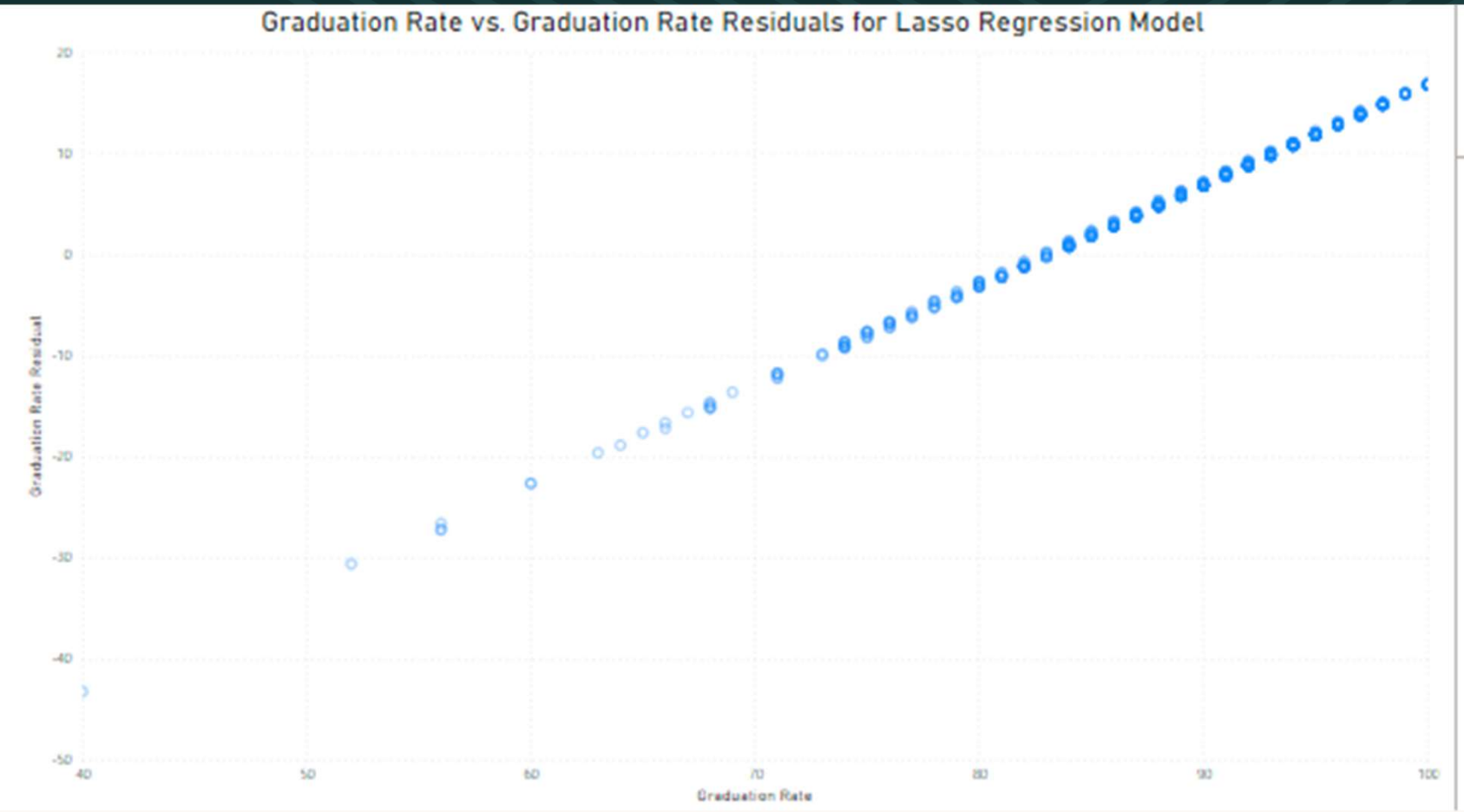


> 60%



Machine Learning

Can we predict the graduation rate for a district based on the data and which factor has the most effect on the graduation rate for a school district?



Filters

Search

There aren't any filters to display.

LassoMLModel, Page 1

Data updated on 8/17/22, 12:32 PM

Highest Effect by Linear Coefficients

- Percentage of Minority Students: -16.7%
- Household Income: 0.016%
- Funding Per Student: -0.011%

Recommendations

- Distribute funds more evenly amongst school districts
- Districts with majority minority student bodies should consider ways of increasing diversity
- Research to discover the educational differences between rural and urban areas
- Further research into the differences between federally funded and non-federally funded districts
- Look at data from years before and after the 2019-2020 school year
    - To understand Covid-19 impacts, and see if there are any lasting trends

PowerBi Dashboard

https://app.powerbi.com/links/OxMVhG-D8y?ctid=d4e104e3-ae7d-4371-a239-745aa8960cc9&pbi_source=linkShare

THANKS

Do you have any questions?

RESOURCES

Create Shape Map Visualizations in Power BI Desktop (preview). Use Shape maps in Power BI Desktop (Preview) - Power BI | Microsoft Docs. (2022). Retrieved from https://docs.microsoft.com/en-us/power-bi/visuals/desktop-shape-map

Department of Health and Mental Hygiene (DOHMH). (2020, May 13). *Modified ZIP code tabulation areas (MODZCTA)*. Modified Zip Code Tabulation Areas (MODZCTA) | NYC Open Data. From https://data.cityofnewyork.us/Health/Modified-Zip-Code-Tabulation-Areas-MODZCTA-/pri4-ifjk

NCES. (2022). *School District Geographic Relationship Files*. Geographic. Retrieved from https://nces.ed.gov/programs/edge/Geographic/RelationshipFiles "\GRF21\grf21_lea_zcta5ce20.xlsx"

New York State. (2019). *2019 District Data – School Funding Transparency*. https://openbudget.ny.gov/sft/sft-districts-19.html

New York State. (2022). *GIS.NY.GOV*. NYS GIS Clearinghouse - NYS Education Department - NYS Schools and School District Boundaries From https://gis.ny.gov/gisdata/inventories/details.cfm?DSID=1326

NYC Open Data. (2022). *School districts*. NYC Open Data. From https://data.cityofnewyork.us/Education/School-Districts/r8nu-ymqj

NYC OpenData. (2021). *FY2020 Local Law 16 Final Report*. https://data.cityofnewyork.us/Education/FY2020-Local-Law-16-Final-Report/cvqn-xqrr/data

NYSED. (2020). *Districts: NYSED Data Site*. data.nysed.gov. From https://data.nysed.gov/lists.php?type=district

U.S. Census Bureau. (2020). *Income in the Past 12 Months (In 2020 Inflation-Adjusted Dollars)*. https://data.census.gov/cedsci/table?q=median+income&g=860XX00US11701

Sielke, C. by C. C. (2001, March 19). *Public School Finance Programs of the U.S. and Canada: 1998-99*. National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education. Retrieved August 17, 2022, from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001309

NYC Class Size 2020. New York City Department of Education. (2020, February). Retrieved August 17, 2022, from https://infohub.nyced.org/docs/default-source/default-document-library/2019-20-february-class-size-report---webdeck---accessible.pdf